

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-345241
 (43)Date of publication of application : 14.12.1999

(51)Int.Cl.

G06F 17/30

(21)Application number : 10-153470

(71)Applicant : FUJITSU LTD

(22)Date of filing : 03.06.1998

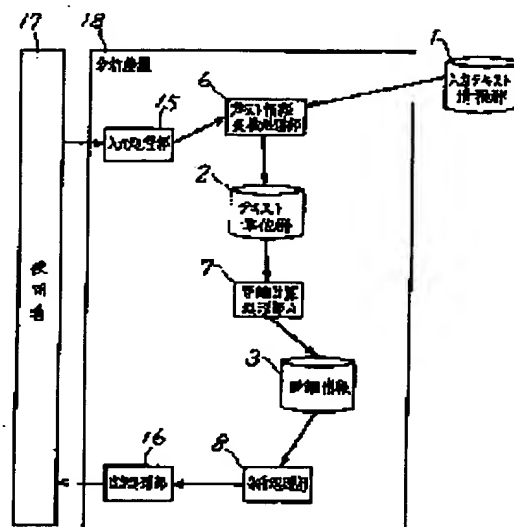
(72)Inventor : WATABE ISAMU

(54) ANALYZER FOR TEXT INFORMATION AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To easily analyze input text information from various angles by extracting a word from a produced text unit group, calculating one distance between the extracted word and a text unit, between words or between text units and performing analysis based on the distance information.

SOLUTION: A user 17 designates a text unit to an input processing part 15. By this, a text information conversion processing part 6 produces a text unit group 2 by performing division, consolidation, working, etc., of an input text information group 1 by a method designated by the part 15. Next, a distance calculation processing part A7 extracts a word from the group 2, calculates distances between a word and a text unit, between words and between text units and produces distance information 3. And, an analysis processing part 8 performs analysis based on the information 3 and an output processing part 16 outputs an analytical result to the user 17.



LEGAL STATUS

[Date of request for examination]

18.05.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C): 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平11-345241

(43)公開日 平成11年(1999)12月14日

(51)Int.Cl.⁶

G 0 6 F 17/30

識別記号

F I

G 0 6 F 15/40
15/401

3 7 0 A
3 2 0 Z

審査請求 未請求 請求項の数6 O L (全 13 頁)

(21)出願番号

特願平10-153470

(22)出願日

平成10年(1998)6月3日

(71)出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72)発明者 渡部 勇

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(74)代理人 弁理士 山谷 皓榮 (外2名)

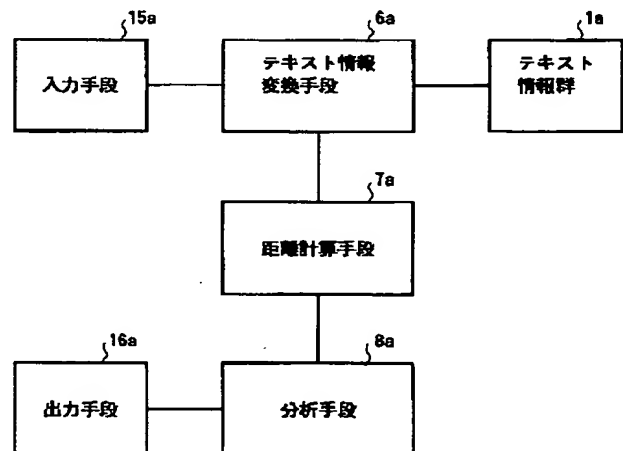
(54)【発明の名称】 テキスト情報の分析装置及び記録媒体

(57)【要約】

【課題】柔軟な目的に応じて、入力テキスト情報をさまざまな角度から容易に分析を行えるようにすること。

【解決手段】テキスト情報群1aより指定された分析の単位となるテキスト単位群を作成するテキスト情報変換手段6aと、該作成したテキスト単位群から単語を抽出し、該抽出した単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する距離計算手段7aと、該計算した距離情報をもとに分析を行う分析手段8aとを備える。

本発明の原理説明図



1

【特許請求の範囲】

【請求項 1】テキスト情報群より指定された分析の単位となるテキスト単位群を作成するテキスト情報変換手段と、

該作成したテキスト単位群から単語を抽出し、該抽出した単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する距離計算手段と、

該計算した距離情報をもとに分析を行う分析手段とを備えることを特徴としたテキスト情報の分析装置。

【請求項 2】前記テキスト単位群に、指定された方法で名前を付ける名前付与手段と、

指定された名前のパターンに合致するものだけを分析の対象にする分析対象限定手段とを備えることを特徴とした請求項 1 記載のテキスト情報の分析装置。

【請求項 3】前記テキスト単位群に、指定された方法で属性値を付ける属性値付与手段と、

指定された属性値を持つものだけを分析の対象にする分析対象限定手段とを備えることを特徴とした請求項 1 記載のテキスト情報の分析装置。

【請求項 4】前記テキスト単位群に付けられた名前のパターンから、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する距離計算手段とを備えることを特徴とした請求項 2 記載のテキスト情報の分析装置。

【請求項 5】前記テキスト単位群に付けられた属性値のパターンから、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する距離計算手段とを備えることを特徴とした請求項 3 記載のテキスト情報の分析装置。

【請求項 6】テキスト情報群より指定された分析の単位となるテキスト単位群を作成するテキスト情報変換手段と、

該作成したテキスト単位群から単語を抽出し、該抽出した単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する距離計算手段と、

該計算した距離情報をもとに分析を行う分析手段として、

コンピュータを機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、目的に応じて、入力テキスト情報をさまざまな角度から分析することができるテキスト情報の分析装置及び記録媒体に関する。

【0002】近年、記憶媒体の大容量化・低価格化、インターネットの急速な普及などにもなって、多量のテキスト情報を計算機を用いて簡単に収集・蓄積することが可能になってきている。

2

【0003】このように集められた情報から何らかの知見を得るような分析を行うためには、その量が膨大であるがゆえに、テキスト情報やその中に現れる単語の間の関係を図解化する機能、テキスト情報や単語をクラスタリング・分類する機能などを持ったテキスト情報の分析装置が不可欠となる。

【0004】テキスト情報の分析装置を用いることにより、例えば、アンケート分析においては、選択式の回答結果だけでなく、自由に記述されたテキスト情報の部分 10 に関しても、統計解析・多変量解析などの数量的な分析手法を利用することが可能となり、多量のデータを用いた精度の高い分析を行うことが可能になる。

【0005】

【従来の技術】現在では、入力テキスト情報群から形態素解析等を用いて単語を抽出する技術、単語の統計的情報（出現頻度等）を利用することにより単語のテキスト情報間の距離を計算する技術などが確立している。

【0006】いったん距離情報が得られれば、あとは従来数量的なデータに対して用いられてきた統計解析・多変量解析などの分析手法を適用することが可能となる。 20 したがって、従来の既知の技術を用いることによって、テキスト情報の分析情報を構成することが可能となる。

【0007】しかし、上記のような従来技術で構成された分析装置では、もともと文書の検索を行うことを目的として開発された技術を用いているため、柔軟な分析を目的とする場合には、テキスト情報の扱いが固定的であることに起因する、次の問題が生じていた。

【0008】（1）：第 1 の問題は、テキスト情報の単位に関するものである。検索装置では、検索の出力して 30 得られるテキスト情報がそのまま結果となるため、例えば論文、特許明細書、あるいはそれを構成する章、節、段落といったように、テキスト情報がそれ自体で意味的なまとまりを持った単位になっていることが必要となる。

【0009】一方、分析装置では、例えばテキスト情報間の関係が結果となるため、句、単語といったように、検索情報ではまとまった情報としてみなされないような単位のテキスト情報を対象にすることも考えられる。

【0010】すなわち、検索装置では容易に想定することができていたテキスト情報のまとまりの単位が、分析装置においては、一般的に明確な基準で決めることができず、むしろ、分析の目的によってその都度変えることができるようになっていくことが必要とされる。

【0011】しかし、従来の技術では、このような目的に応じて入力テキスト情報のまとまりの単位を変更するための簡便な手段が提供されておらず、あらかじめ前処理によって、入力テキスト情報を分割、統合、加工し、 40 適当なレベルのテキスト単位に変換しておく必要があった。

50 【0012】例えば、複数人に対して、複数の設問を行

3

い、各設問には複数の回答欄があり、各回答欄には複数の文を書くことができるようなアンケート結果の分析を行う場合なら、分析の目的によって、各設問に対する全員の回答をまとめて一つの単位とする、一人分の回答をまとめて一つの単位とする、各回答者の各設問の回答を一つの単位とする、それを更に回答欄別に扱う、といったテキスト情報の単位の取り方のバリエーションが考えられる。しかし、それぞれのバリエーションごとに予めアンケート分析結果の回答テキスト情報を分割、統合、加工した文書群のセットを用意しておくことが必要となっていた。

【0013】(2)：第2の問題は、分析の対象となるテキスト情報の制限方法に関するものである。テキスト情報の検索では、例えば、特許明細書の検索において、特定の年に出版されたものだけを検索する、IPC（国際特許分類）の情報を利用して特定の分野のものだけを検索するといったように、目的に応じて対象とするテキスト情報群を制限する機能が有効である場合がある。

【0014】テキスト情報の分析においても、検索の場合と同様に、対象を制限することが有効である場合が考えられる。検索の場合に、対象を制限することができるのは、対象となるテキスト情報に何らかの属性が付与されていて、それを用いることが可能だからである。

【0015】しかし、分析の場合には、前述したように、そもそも属性を付与すべき対象であるテキスト情報の単位が明確でなかった。例えば、前述のアンケート分析の例において、各回答者の年齢が分かっている場合、一人分の回答をまとめて一つの単位とするケースでは、各テキスト情報の単位の属性として回答者の年齢を付与することができるが、各設問に対する全員の回答をまとめて一つの単位とするケースでは、そのような属性を付与することは意味を持たない。

【0016】また、入力テキスト情報群に属性が付与されていたとしても、分割、統合、加工の際に、その属性情報もまた適切な変換手段を用いて付与し直す作業が別途発生していた。

【0017】(3)：第3の問題は、入力テキスト情報を分割することによって、もともと持っていた情報を失ってしまうことに関するものである。例えば、前述のアンケート分析の例において、入力テキスト情報を、各回答者の各設問の各回答欄の回答を分析の単位として扱うように分割した場合、同一回答者の同一設問に対する回答の間には強い関連がある、同一設問に対する回答の間には何らかの関連があるといった、本来入力テキスト情報が持っていた情報を失ってしまうことになっていた。

【0018】

【発明が解決しようとする課題】前記従来のものにおいては、次のような課題があった。従来技術を用いたテキスト情報検索装置では、柔軟な目的に対応するためには、入力テキスト情報群を、前処理によって分割、統

4

合、加工したりする作業が必要であった。

【0019】本発明は、このような従来の課題を解決し、柔軟な目的に応じて、入力テキスト情報をさまざまな角度から容易に分析を行えるようにすることを目的とする。

【0020】

【課題を解決するための手段】図1は本発明の原理説明図である。図1中、1aはテキスト情報群、6aはテキスト情報変換手段、7aは距離計算手段、8aは分析手段、15aは入力手段、16aは出力手段である。

【0021】本発明は前記従来の課題を解決するため次のように構成した。

(1)：テキスト情報群1aより指定された分析の単位となるテキスト単位群を作成するテキスト情報変換手段6aと、該作成したテキスト単位群から単語を抽出し、該抽出した単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する距離計算手段7aと、該計算した距離情報をもとに分析を行う分析手段8aとを備える。

【0022】(2)：前記(1)のテキスト情報の分析装置において、前記テキスト単位群に、指定された方法で名前を付ける名前付与手段と、指定された名前のパターンに合致するものだけを分析の対象にする分析対象限定手段とを備える。

【0023】(3)：前記(1)のテキスト情報の分析装置において、前記テキスト単位群に、指定された方法で属性値を付ける属性値付与手段と、指定された属性値を持つものだけを分析の対象にする分析対象限定手段とを備える。

【0024】(4)：前記(2)のテキスト情報の分析装置において、前記テキスト単位群に付けられた名前のパターンから、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する距離計算手段とを備える。

【0025】(5)：前記(3)のテキスト情報の分析装置において、前記テキスト単位群に付けられた属性値のパターンから、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する距離計算手段とを備える。

【0026】(6)：テキスト情報群1aより指定された分析の単位となるテキスト単位群を作成するテキスト情報変換手段6aと、該作成したテキスト単位群から単語を抽出し、該抽出した単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する距離計算手段7aと、該計算した距離情報をもとに分析を行う分析手段8aとして、コンピュータを機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体とする。

【0027】(作用)前記構成に基づく作用を説明する。本発明は前記従来の課題を解決するため次のように

構成した。

【0028】テキスト情報変換手段6aでテキスト情報群1aより指定された分析の単位となるテキスト単位群を作成し、距離計算手段7aで該作成したテキスト単位群から単語を抽出し、該抽出した単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算し、分析手段8aで該計算した距離情報をもとに分析を行う。このため、柔軟な目的に応じて、入力テキスト情報をさまざまな角度から容易に分析を行える。

【0029】また、名前付与手段で前記テキスト単位群に、指定された方法で名前を付け、分析対象限定手段で指定された名前のパターンに合致するものだけを分析の対象にする。このため、テキスト単位の名前にテキスト情報群の属性情報を埋め込むような指定を行っておけば、名前のパターンを指定することにより、特別な前処理なしに、目的に応じて分析対象とするテキスト情報群を制限することができる。

【0030】さらに、属性値付与手段で前記テキスト単位群に、指定された方法で属性値を付け、分析対象限定手段で指定された属性値を持つものだけを分析の対象にする。このため、特別な前処理なしに、目的に応じて分析対象とするテキスト情報群を制限することができる。

【0031】また、距離計算手段で前記テキスト単位群に付けられた名前のパターンから、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する。このため、入力テキスト情報群の中で関連のあったテキスト単位群に、共通する名前を付けることによって、もともと持っていた関連性の情報を失うことなく分析を行うことができる。

【0032】さらに、距離計算手段で前記テキスト単位群に付けられた属性値のパターンから、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する。このため、入力テキスト情報群の中で関連のあったテキスト単位群に、共通する属性値を付けることによって、もともと持っていた関連性の情報を失うことなく分析を行うことができる。

【0033】また、テキスト情報群1aより指定された分析の単位となるテキスト単位群を作成するテキスト情報変換手段6aと、該作成したテキスト単位群から単語を抽出し、該抽出した単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する距離計算手段7aと、該計算した距離情報をもとに分析を行う分析手段8aとして、コンピュータを機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体とする。このため、この記録媒体のプログラムをコンピュータにインストールすることで、柔軟な目的に応じて、入力テキスト

情報群をさまざまな角度から容易に分析を行えるテキスト情報の分析装置を容易に提供することができる。

【0034】

【発明の実施の形態】以下、図面に基づいて本発明の実施の形態について説明する。

(1)：テキスト単位群を自動作成する場合の説明

A：テキスト単位群を自動作成する分析装置の説明

図2はテキスト単位群を自動作成する分析装置の説明図である。図2において、分析装置18は、使用者17から指定された方法で入力テキスト情報群1を分析して出力するものである。

【0035】分析装置18には、テキスト単位群2、距離情報3、テキスト情報変換処理部6、距離計算処理部A7、分析処理部8、入力処理部15、出力処理部16が設けてある。

【0036】テキスト単位群2は、テキスト情報変換処理部6が入力テキスト情報群1を指定された方法で分割、統合、加工等を行って作成したものである。距離情報3は、距離計算処理部A7でテキスト単位群2から単語を抽出し、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離を計算したものである。テキスト情報変換処理部6は、入力テキスト情報群1を入力処理部15から指定された方法で分割、統合、加工等を行ってテキスト単位群2を得るものである。距離計算処理部A7は、テキスト単位群2から単語を抽出し、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離情報3を計算するものである。分析処理部8は、距離情報3をもとに分析を行うものである。入力処理部15は、使用者17が行う入力を処理するものである。出力処理部16は、分析結果を使用者17に出力するものである。

【0037】(動作の説明)使用者17は、入力処理部15にテキスト単位の指定を行う。これにより、テキスト情報変換処理部6で、入力テキスト情報群1を入力処理部15から指定された方法で分割、統合、加工等を行ってテキスト単位群2が作成される。次に、距離計算処理部A7で、テキスト単位群2から単語を抽出し、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離を計算し距離情報3を作成する。そして、分析処理部8で、距離情報3をもとに分析を行い、出力処理部16で、分析結果を使用者17に出力するのである。

【0038】B：テキスト単位群を自動作成するシステム構成の説明

図3はテキスト単位群を自動作成するシステム構成の説明図である。図3において、データ処理装置19は、外部記憶装置20、21、22、入力装置25、出力装置26と接続されている。データ処理装置19には、テキスト情報変換処理部6、距離計算処理部A7、分析処理部8、入力処理部15、出力処理部16が設けてある。

【0039】テキスト情報変換処理部6は、入力テキスト情報群1を入力処理部15から指定された方法で分割、統合、加工等を行ってテキスト単位群2を得るものである。距離計算処理部A7は、テキスト単位群2から単語を抽出し、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離情報3を計算するものである。分析処理部8は、距離情報3をもとに分析を行うものである。入力処理部15は、使用者17が入力装置25から行う入力を処理するものである。出力処理部16は、分析結果を出力装置26により使用者17に出力するものである。

【0040】データ処理装置19は、CPU（中央処理装置）とメモリ等から構成されるものである。外部記憶装置20は、テキスト情報変換処理部6で使用する入力テキスト情報群1を格納するものである。外部記憶装置21は、テキスト情報変換処理部6によって得られるテキスト単位群2を格納するものである。外部記憶装置22は、距離計算処理部A7によって得られる距離情報3を格納するものである。

【0041】入力装置25は、キーボード等から構成され、使用者17からの入力を入力処理部15に伝えるものである。出力装置26は、ディスプレイ装置等から構成され、出力処理部16から得られる出力を使用者17に表示するものである。

【0042】（動作の説明）入力装置25から入力処理部15にテキスト単位の指定を行う。これにより、テキスト情報変換処理部6で、外部記憶装置20の入力テキスト情報群1を入力処理部15から指定された方法で分割、統合、加工等を行ってテキスト単位群2を作成し、外部記憶装置21に格納する。次に、距離計算処理部A7で、外部記憶装置21のテキスト単位群2から単語を抽出し、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離を計算して距離情報3を作成し、外部記憶装置22に格納する。そして、分析処理部8で、外部記憶装置22の距離情報3をもとに分析を行い、出力処理部16で、分析結果を出力装置26に出力するものである。

【0043】（2）：テキスト単位群に名前を付ける場合の説明

A：テキスト単位群に名前を付ける分析装置の説明

図4はテキスト単位群に名前を付ける分析装置の説明図である。図4において、分析装置18は、図2の構成に、名前情報4、名前付与処理部9、分析対象限定処理部A11を追加したものである。

【0044】名前情報4は、名前付与処理部9でテキスト単位群2に付けるもので、パターンマッチが可能な規則的な名前である。名前付与処理部9は、テキスト単位群2を入力処理部15から指定された方法で規則的な名前を付けるものである。分析対象限定処理部A11は、入力処理部15から指定された名前のパターンに合致す

るものだけを分析の対象にするものである。

【0045】（動作の説明）使用者17は、入力処理部15にテキスト単位の指定を行う。これにより、テキスト情報変換処理部6で、入力テキスト情報群1を入力処理部15から指定された方法で分割、統合、加工等を行ってテキスト単位群2が作成される。次に、名前付与処理部9でテキスト単位群2を入力処理部15から指定された方法で規則的な名前を付け名前情報4を作成する。距離計算処理部A7では、テキスト単位群2から単語を抽出し、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離を計算し距離情報3を作成する。さらに、分析対象限定処理部A11で入力処理部15から指定された名前のパターンに合致するものだけを分析の対象にするように、テキスト単位群2を制限する。そして、分析処理部8で、距離情報3をもとに分析を行い、出力処理部16で、分析結果を使用者17に出力するものである。

【0046】B：テキスト単位群に名前を付けるシステム構成の説明

図5はテキスト単位群に名前を付けるシステム構成の説明図である。図5において、システム構成は、図3の構成に、名前付与処理部9、分析対象限定処理部A11、外部記憶装置23を追加したものである。

【0047】名前付与処理部9は、テキスト単位群2を入力処理部15から指定された方法で規則的な名前を付けるものである。分析対象限定処理部A11は、入力処理部15から指定された名前のパターンに合致するものだけを分析の対象にするものである。外部記憶装置23は、名前付与処理部9でテキスト単位群2を入力処理部15から指定された方法で付けた規則的な名前情報を格納するものである。

【0048】（動作の説明）入力装置25から入力処理部15にテキスト単位の指定を行う。これにより、テキスト情報変換処理部6で、外部記憶装置20の入力テキスト情報群1を入力処理部15から指定された方法で分割、統合、加工等を行ってテキスト単位群2を作成し、外部記憶装置21に格納する。次に、名前付与処理部9でテキスト単位群2を入力処理部15から指定された方法で規則的な名前を付け名前情報4を作成し、外部記憶装置23に格納する。距離計算処理部A7では、テキスト単位群2から単語を抽出し、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離を計算して距離情報3を作成し、外部記憶装置22に格納する。さらに、分析対象限定処理部A11で入力処理部15から指定された名前のパターンに合致するものだけを分析の対象にするように、テキスト単位群2を制限する。そして、分析処理部8で、距離情報3をもとに分析を行い、出力処理部16で、分析結果を出力装置26に出力するものである。

【0049】以上のように、テキスト単位群2に指定さ

れた方法で規則的な名前情報 4 を付ける名前付与処理部 9、及び、指定された名前のパターンに合致するものだけを分析の対象にする分析対象限定処理部 A 11 を備えているので、テキスト単位の名前にテキスト情報群の属性情報を埋め込むような指定を行っておけば、名前のパターンを指定することにより、特別な前処理なしに、目的に応じて分析対象とするテキスト情報群を制限することができる。

【0050】(3)：テキスト単位群に属性値を付ける場合の説明

A：テキスト単位群に属性値を付ける分析装置の説明
図 6 はテキスト単位群に属性値を付ける分析装置の説明図である。図 6 において、分析装置 18 は、図 4 の構成における、名前情報 4 が属性値情報 5 に、名前付与処理部 9 が属性値付与処理部 10 に、分析対象限定処理部 A 11 が分析対象限定処理部 B 12 に、それぞれ置き換わったものである。

【0051】属性値情報 5 は、属性値付与処理部 10 でテキスト単位群 2 に指定された方法で付与された属性値である。属性値付与処理部 10 は、テキスト単位群 2 に入力処理部 15 から指定された方法で属性値を付与するものである。分析対象限定処理部 B 12 は、入力処理部 15 から指定された属性値を持つものだけを分析の対象にするものである。

【0052】(動作の説明) 使用者 17 は、入力処理部 15 にテキスト単位の指定を行う。これにより、テキスト情報変換処理部 6 で、入力テキスト情報群 1 を入力処理部 15 から指定された方法で分割、統合、加工等を行ってテキスト単位群 2 が作成される。次に、属性値付与処理部 10 でテキスト単位群 2 に入力処理部 15 から指定された方法で属性値を付け属性値情報 5 を作成する。距離計算処理部 A 7 では、テキスト単位群 2 から単語を抽出し、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離を計算し距離情報 3 を作成する。さらに、分析対象限定処理部 B 12 で入力処理部 15 から指定された属性値を持つものだけを分析の対象にするように、テキスト単位群 2 を制限する。そして、分析処理部 8 で、距離情報 3 をもとに分析を行い、出力処理部 16 で、分析結果を使用者 17 に出力するものである。

【0053】B：テキスト単位群に属性値を付けるシステム構成の説明

図 7 はテキスト単位群に属性値を付けるシステム構成の説明図である。図 7 において、システム構成は、図 3 の構成に、属性値付与処理部 10、分析対象限定処理部 B 12、外部記憶装置 24 を追加したものである。

【0054】属性値付与処理部 10 は、テキスト単位群 2 に入力処理部 15 から指定された方法で属性値を付与するものである。分析対象限定処理部 B 12 は、入力処理部 15 から指定された属性値に合致するものだけを分

析の対象にするものである。外部記憶装置 24 は、属性値付与処理部 10 でテキスト単位群 2 に入力処理部 15 から指定された方法で付与した属性値情報 5 を格納するものである。

【0055】(動作の説明) 入力装置 25 から入力処理部 15 にテキスト単位の指定を行う。これにより、テキスト情報変換処理部 6 で、外部記憶装置 20 の入力テキスト情報群 1 を入力処理部 15 から指定された方法で分割、統合、加工等を行ってテキスト単位群 2 を作成し、外部記憶装置 21 に格納する。次に、属性値付与処理部 10 でテキスト単位群 2 に入力処理部 15 から指定された方法で属性値を付け属性値情報 5 を作成し、外部記憶装置 24 に格納する。距離計算処理部 A 7 で、テキスト単位群 2 から単語を抽出し、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離を計算して距離情報 3 を作成し、外部記憶装置 22 に格納する。さらに、分析対象限定処理部 B 12 で入力処理部 15 から指定された属性値を持つものだけを分析の対象にするように、テキスト単位群 2 を制限する。そして、分析処理部 8 で、距離情報 3 をもとに分析を行い、出力処理部 16 で、分析結果を出力装置 26 に出力するものである。

【0056】以上のように、テキスト単位群 2 に指定された方法で属性値情報 5 を付与する属性値付与処理部 10、及び、指定された属性値を持つものだけを分析の対象にする分析対象限定処理部 B 12 を備えているので、特別な前処理なしに、目的に応じて分析対象とするテキスト情報群を制限することができる。

【0057】(4)：テキスト単位群の名前のパターンから距離を計算する場合の説明

A：テキスト単位群の名前のパターンから距離を計算する分析装置の説明

図 8 はテキスト単位群の名前のパターンから距離を計算する分析装置の説明図である。図 8 において、分析装置 18 は、図 4 の構成に、距離計算処理部 B 13 を追加したものである。

【0058】距離計算処理部 B 13 は、テキスト単位群 2 に付けられた名前情報 4 のパターンから単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離情報 3 を計算するものである。

【0059】(動作の説明) 使用者 17 は、入力処理部 15 にテキスト単位の指定を行う。これにより、テキスト情報変換処理部 6 で、入力テキスト情報群 1 を入力処理部 15 から指定された方法で分割、統合、加工等を行ってテキスト単位群 2 が作成される。次に、名前付与処理部 9 でテキスト単位群 2 に入力処理部 15 から指定された方法で規則的な名前を付け名前情報 4 を作成する。距離計算処理部 A 7 で、テキスト単位群 2 から単語を抽出し、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離を計算し距離情報 3 を作成

する。さらに、距離計算処理部B13で、名前情報4の
パターンから単語・テキスト単位間、単語・単語間、テ
キスト単位・テキスト単位間の距離を計算し距離情報3
を作成する。そして、分析処理部8で、距離情報3をも
とに分析を行い、出力処理部16で、分析結果を使用者
17に出力するものである。

【0060】B：テキスト単位群の名前のパターンから
距離を計算するシステム構成の説明

図9はテキスト単位群の名前のパターンから距離を計算
するシステム構成の説明図である。図9において、デー
タ処理装置19は、図5の構成に、距離計算処理部B1
3を追加したものである。

【0061】距離計算処理部B13は、テキスト単位群
2に付けられた名前情報4のパターンから単語・テキス
ト単位間、単語・単語間、テキスト単位・テキスト単位
間の距離情報3を計算するものである。

【0062】（動作の説明）入力装置25から入力処理
部15にテキスト単位の指定を行う。これにより、テキ
スト情報変換処理部6で、外部記憶装置20の入力テキ
スト情報群1を入力処理部15から指定された方法で分
割、統合、加工等を行ってテキスト単位群2を作成し、
外部記憶装置21に格納する。次に、名前付与処理部9
でテキスト単位群2を入力処理部15から指定された方
法で規則的な名前を付け名前情報4を作成し、外部記憶
装置23に格納する。距離計算処理部A7で、テキスト
単位群2から単語を抽出し、単語・テキスト単位間、単
語・単語間、テキスト単位・テキスト単位間の距離を計
算して距離情報3を作成し、外部記憶装置22に格納す
る。さらに、距離計算処理部B13で、名前情報4のパ
ターンから単語・テキスト単位間、単語・単語間、テキ
スト単位・テキスト単位間の距離を計算して距離情報3
を作成し、外部記憶装置22に格納する。そして、分析
処理部8で、距離情報3をもとに分析を行い、出力処理
部16で、分析結果を出力装置26に出力するものでは
ある。

【0063】以上のように、テキスト単位群2に付けら
れた名前情報4のパターンから単語・テキスト単位間、
単語・単語間、テキスト単位・テキスト単位間の距離情
報3を計算する距離計算処理部B13を備えているの
で、入力テキスト情報群の中で関連のあったテキスト単
位群に、共通する名前を付けることによって、もともと
持っていた関連性の情報を失うことなく分析を行うこと
ができる。

【0064】（5）：テキスト単位群の属性値のパター
ンから距離を計算する場合の説明

A：属性値のパターンから距離を計算する分析装置の説
明

図10はテキスト単位群の属性値のパターンから距離を
計算する分析装置の説明図である。図10において、分
析装置18は、図8の構成における、名前情報4が属性

値情報5に、名前付与処理部9が属性値付与処理部10
に、距離計算処理部B13が距離計算処理部C14に、
それぞれ置き換わったものである。

【0065】属性値情報5は、属性値付与処理部10で
テキスト単位群2に指定された方法で付与された属性値
である。属性値付与処理部10は、テキスト単位群2に
入力処理部15から指定された方法で属性値を付与する
ものである。距離計算処理部C14は、テキスト単位群
2に付けられた属性値情報5のパターンから単語・テキ
スト単位間、単語・単語間、テキスト単位・テキスト単
位間の距離情報3を計算するものである。

【0066】（動作の説明）使用者17は、入力処理部
15にテキスト単位の指定を行う。これにより、テキス
ト情報変換処理部6で、入力テキスト情報群1を入力処
理部15から指定された方法で分割、統合、加工等を行
ってテキスト単位群2が作成される。次に、属性値付与
処理部10でテキスト単位群2を入力処理部15（使用
者17）から指定された方法で属性値を付け属性値情報
5を作成する。距離計算処理部A7で、テキスト単位群
2から単語を抽出し、単語・テキスト単位間、単語・単
語間、テキスト単位・テキスト単位間の距離を計算し距
離情報3を作成する。さらに、距離計算処理部C14
で、属性値情報5のパターンから単語・テキスト単位
間、単語・単語間、テキスト単位・テキスト単位間の距
離を計算し距離情報3を作成する。そして、分析処理部
8で、距離情報3をもとに分析を行い、出力処理部16
で、分析結果を使用者17に出力するものである。

【0067】B：属性値のパターンから距離を計算する
システム構成の説明

図11はテキスト単位群の属性値のパターンから距離を
計算するシステム構成の説明図である。図11におい
て、データ処理装置19は、図7の構成に、距離計算処
理部C14を追加したものである。

【0068】距離計算処理部C14は、テキスト単位群
2に付けられた属性値情報5のパターンから単語・テキ
スト単位間、単語・単語間、テキスト単位・テキスト単
位間の距離情報3を計算するものである。

【0069】（動作の説明）入力装置25から入力処理
部15にテキスト単位の指定を行う。これにより、テキ
スト情報変換処理部6で、外部記憶装置20の入力テキ
スト情報群1を入力処理部15から指定された方法で分
割、統合、加工等を行ってテキスト単位群2を作成し、
外部記憶装置21に格納する。次に、属性値付与処理部
10でテキスト単位群2を入力処理部15（入力装置2
5）から指定された方法で属性値を付け属性値情報5を
作成し、外部記憶装置23に格納する。距離計算処理部
A7で、テキスト単位群2から単語を抽出し、単語・テ
キスト単位間、単語・単語間、テキスト単位・テキスト
単位間の距離を計算し距離情報3を作成し、外部記憶装
置22に格納する。さらに、距離計算処理部C14で、

属性値情報5のパターンから単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離を計算した距離情報3を作成し、外部記憶装置22に格納する。そして、分析処理部8で、距離情報3をもとに分析を行い、出力処理部16で、分析結果を出力装置26に出力するものである。

【0070】以上のように、テキスト単位群2に付けられた属性値情報5のパターンから単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間の距離情報3を計算する距離計算処理部C14を備えているので、入力テキスト情報群の中で関連のあったテキスト単位群に、共通する属性値を付けることによって、もともと持っていた関連性の情報を失うことなく分析を行うことができる。

【0071】(6)：アンケート結果のテキスト情報を分析する場合の説明

前記説明した属性値付与処理部10は名前付与処理部9と同様の動作を行い、分析対象限定処理部B12は分析対象限定処理部A11と同様の動作を行い、距離計算処理部C14は距離計算処理部B13と同様の動作を行っている。ここでは、テキスト情報変換処理部6、名前付与処理部9、分析対象限定処理部A11、距離計算処理部B13に関し、図12のようなアンケート結果のテキスト情報を分析するケースを例として説明する。

【0072】図12はアンケート結果表の説明図である。図12において、回答者ID、回答者の年齢、回答者の職業、設問1：商品Aのイメージ、設問2：商品Bのイメージ欄が設けてある。設問1：商品Aのイメージ欄には回答欄1、回答欄2、回答欄3が設けてあり、設問2：商品Bのイメージ欄には回答欄1、回答欄2、回答欄3が設けてある。

【0073】・図12に示すように入力テキスト情報が、表形式のデータとして与えられている場合、使用者17は、どの行・列のデータをまとめてテキスト単位とするのかを、テキスト情報変換処理部6に指定し、テキスト情報変換処理部6は、使用者17の指定に従って、行・列のデータからテキスト情報を作成する。

【0074】例えば、一人の全回答を一つのテキスト単位とする場合には、各行について、4列から9列までのデータ（設問1と設問2の回答）をつなげたテキスト情報を作成するように指定する。

【0075】また、各回答者の各回答欄のデータを一つのテキスト単位とする場合には、各行について、4列から9列までのデータをそれぞれ一つのテキスト情報とするように指定する。

【0076】テキスト単位を作成する際に、単に行・列のデータをつなげるだけでなく、指定した文字列を付加するように指定することもできる。このようにすれば、例えば、4列目から6列目までのデータには、設問に含まれる「商品A」という文字列を付加（例えば、商品名

の一部である「A」のパターンを付加)し、7列目から9列目までのデータには、設問に含まれる「商品B」という文字列を付加することによって、各テキスト単位がどの設問に対する回答であるかという情報を、テキスト単位の中に埋め込むことが可能となる。

【0077】さらに、テキスト単位を作成する際に、一つのデータを指定した文字で分割するように指定することもできる。このようにすれば、例えば「。」という文字で分割するように指定すれば、各回答欄に書かれている一つの文をテキスト単位として設定することもできるようになる。

【0078】・テキスト単位に対する名前は、使用者17が、行番号や列番号や各行・列に含まれるデータといった情報の組み合わせ方・書式などを、名前付与処理部9に指定し、名前付与処理部9が作成するものである。

【0079】例えば、各回答者の各回答欄のデータを一つのテキスト単位とする場合であれば、1列目のデータ、2列目のデータ、3列目のデータおよび列番号を組み合わせて名前を作成するように指定することにより、各テキスト単位の名前に、回答者ID、回答者の年齢、回答者の職業、どの設問の何番目の回答であるかといった情報を埋め込むことができる。

【0080】テキスト単位の名前の一部に、任意の文字列を含むように指定する方法も考えられる。このようにすれば、例えば、4列目から6列目までのデータから作成されるテキスト単位の名前には「商品A」という文字列を含むようにする（例えば、商品名の一部である

「A」等のパターンをハイホン「-」等で区切って含ませる）、7列目から9列目までのデータから作成されるテキスト単位の名前には「商品B」という文字列を含むようにするといったことが可能となり、各テキスト単位の名前に、どの設問に対する回答であるかという情報を埋め込むこともできるようになる。

【0081】・分析対象を限定するために指定する名前のパターンは、正規表現やワイルドカードなどによって指定する。分析対象限定処理部A11は、指定されたパターンにマッチする名前を持つテキスト単位だけを対象とするように限定処理を行う。

【0082】例えば、テキスト単位の名前に埋め込まれた回答者の年齢の部分が、ある特定の範囲の数値にマッチするような正規表現のパターンを指定すれば、特定の年齢の回答者の回答だけを分析対象とするといった限定が可能になる。

【0083】・名前情報4のパターンから単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間等の距離情報3を計算する場合は、テキスト単位の名前を構成するどの要素（部分文字列）が一致していれば距離を定義するのかを指定する。距離計算処理部Bは、各テキスト単位の名前を比較し、指定された部分が一致しているテキスト単位間の距離を指定された方法で再計

算する。

【0084】距離の再計算の方法としては、距離計算処理部Aにより計算される距離を無視して、距離計算処理部Bの値をそのまま使う方法、距離計算処理部Aにより計算される距離に距離計算処理部Bの値を掛ける方法、距離計算処理部Aにより計算される距離と距離計算処理部Bの値を重み付け平均する方法などが考えられる。

【0085】例えば、各回答者の各回答欄のデータを一つのテキスト単位とするケースで、名前のうち、回答者IDの部分と設問を示す文字列（前記の「商品A」「商品B」）の部分が一致している場合には、距離を半分にするという指定をすれば、同一回答者の同一設問に対する回答の間には何らかの関連があるといった情報を加味して、分析を行うことが可能となる。

【0086】以上実施の形態で説明したように、目的に応じて入力テキスト情報を分割、統合、加工して扱ったり、分析の対象を限定したり、入力テキスト情報に内在する関係情報も扱うことができるようにしているので、柔軟な目的に応じて、入力テキスト情報をさまざまな角度から分析することが容易に可能である。このため、アンケート分析などのテキスト情報の分析作業の質の向上、手間の軽減に寄与するところが大きい。

【0087】（7）：プログラムのインストールの説明
テキスト情報変換処理部6、距離計算処理部A7、分析処理部8、名前付与処理部9、属性値付与処理部10、分析対象限定処理部A11、分析対象限定処理部B12、距離計算処理部B13、距離計算処理部C14、入力処理部15、出力処理部16等は、プログラムで構成でき、主制御部（CPU）が実行するものであり、主記憶に格納されているものである。これらのプログラムは、一般的な、パーソナルコンピュータ、ワークステーション等のデータ処理装置（コンピュータ）で処理されるものである。このコンピュータは、主制御部、主記憶、ハードディスク等のファイル装置、表示装置、キーボード等の入力手段である入力装置などのハードウェアで構成されている。

【0088】このコンピュータに、本発明のプログラムをインストールする。このインストールは、フロッピー、光磁気ディスク等の可搬型の記録（記憶）媒体に、これらのプログラムを記憶させておき、コンピュータが備えている記録媒体に対して、アクセスするためのドライブ装置を介して、或いは、LAN等のネットワークを介して、コンピュータに設けられたファイル装置にインストールされる。そして、このファイル装置から処理に必要なプログラムステップを主記憶に読み出し、主制御部が実行するものである。

【0089】

【発明の効果】以上説明したように、本発明によれば次のような効果がある。

（1）：テキスト情報変換手段でテキスト情報群より指

定された分析の単位となるテキスト単位群を作成し、距離計算手段で該作成したテキスト単位群から単語を抽出し、該抽出した単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算し、分析手段で該計算した距離情報をもとに分析を行うため、柔軟な目的に応じて、入力テキスト情報をさまざまな角度から容易に分析を行える。

【0090】（2）：名前付与手段でテキスト単位群に、指定された方法で名前を付け、分析対象限定手段で指定された名前のパターンに合致するものだけを分析の対象にするため、テキスト単位の名前にテキスト情報群の属性情報を埋め込むような指定を行っておけば、名前のパターンを指定することにより、特別な前処理なしに、目的に応じて分析対象とするテキスト情報群を制限することができる。

【0091】（3）：属性値付与手段で前記テキスト単位群に、指定された方法で属性値を付け、分析対象限定手段で指定された属性値を持つものだけを分析の対象にするため、特別な前処理なしに、目的に応じて分析対象とするテキスト情報群を制限することができる。

【0092】（4）：距離計算手段で前記テキスト単位群に付けられた名前のパターンから、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算するため、入力テキスト情報群の中で関連のあったテキスト単位群に、共通する名前を付けることによって、もともと持っていた関連性の情報を失うことなく分析を行うことができる。

【0093】（5）：距離計算手段で前記テキスト単位群に付けられた属性値のパターンから、単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算するため、入力テキスト情報群の中で関連のあったテキスト単位群に、共通する属性値を付けることによって、もともと持っていた関連性の情報を失うことなく分析を行うことができる。

【0094】（6）：テキスト情報群より指定された分析の単位となるテキスト単位群を作成するテキスト情報変換手段と、該作成したテキスト単位群から単語を抽出し、該抽出した単語・テキスト単位間、単語・単語間、テキスト単位・テキスト単位間のうち少なくとも一つの間の距離を計算する距離計算手段と、該計算した距離情報をもとに分析を行う分析手段として、コンピュータを機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体とするため、この記録媒体のプログラムをコンピュータにインストールすることで、柔軟な目的に応じて、入力テキスト情報群をさまざまな角度から容易に分析を行えるテキスト情報の分析装置を容易に提供することができる。

【図面の簡単な説明】

【図1】本発明の原理説明図である。

【図 2】実施の形態におけるテキスト単位群を自動作成する分析装置の説明図である。

【図 3】実施の形態におけるテキスト単位群を自動作成するシステム構成の説明図である。

【図 4】実施の形態におけるテキスト単位群に名前を付ける分析装置の説明図である。

【図 5】実施の形態におけるテキスト単位群に名前を付けるシステム構成の説明図である。

【図 6】実施の形態におけるテキスト単位群に属性値を付ける分析装置の説明図である。

【図 7】実施の形態におけるテキスト単位群に属性値を付けるシステム構成の説明図である。

【図 8】実施の形態におけるテキスト単位群の名前のパターンから距離を計算する分析装置の説明図である。

【図 9】実施の形態におけるテキスト単位群の名前のパターンから距離を計算するシステム構成の説明図であ

る。

【図 1 0】実施の形態におけるテキスト単位群の属性値のパターンから距離を計算する分析装置の説明図である。

【図 1 1】実施の形態におけるテキスト単位群の属性値のパターンから距離を計算するシステム構成の説明図である。

【図 1 2】実施の形態におけるアンケート結果表の説明図である。

10 【符号の説明】

1 a テキスト情報群

6 a テキスト情報変換手段

7 a 距離計算手段

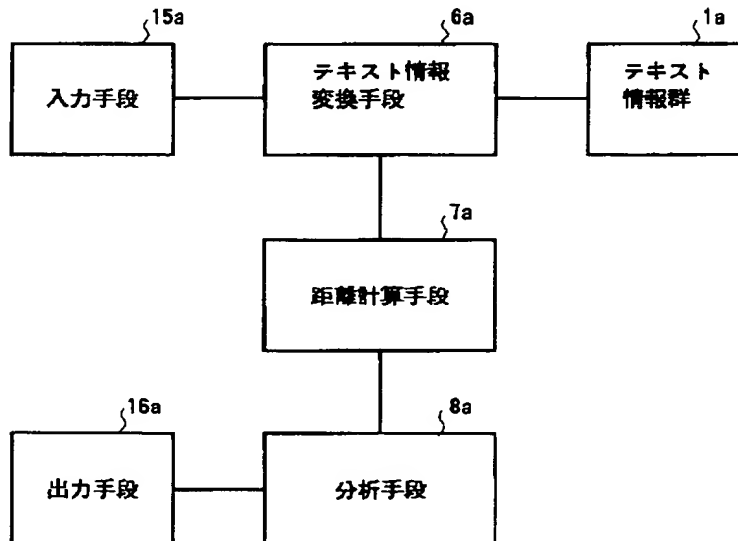
8 a 分析手段

1 5 a 入力手段

1 6 a 出力手段

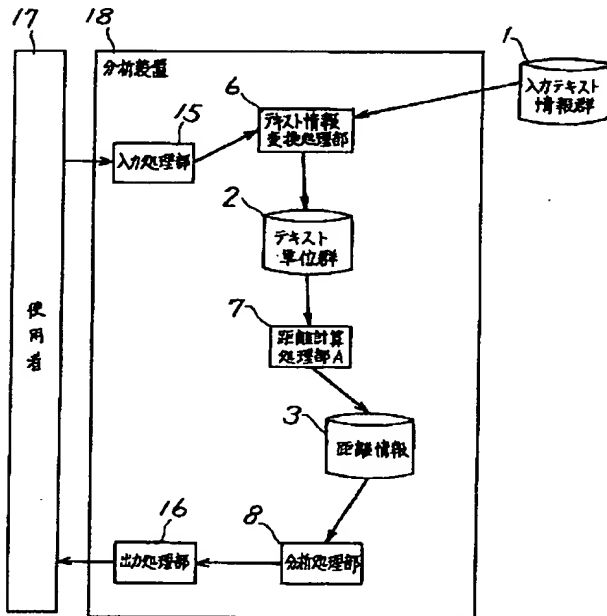
【図 1】

本発明の原理説明図



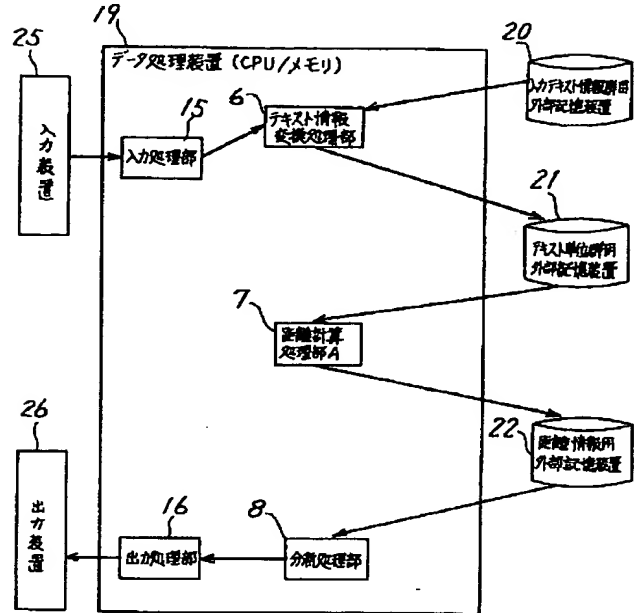
【図 2】

テキスト単位群を自動作成する分析装置の
説明図



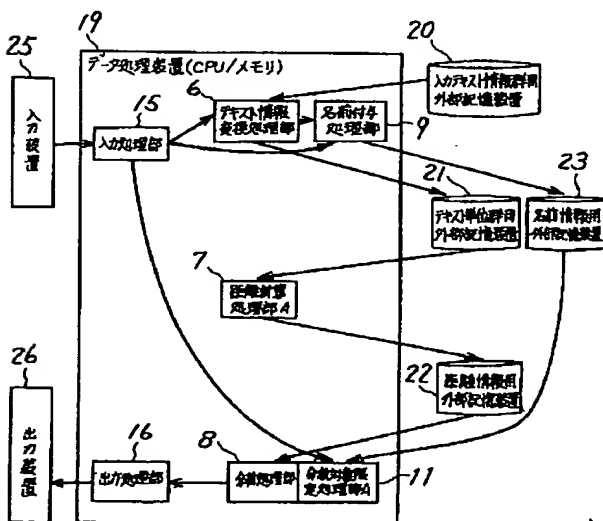
【図 3】

テキスト単位群を自動作成するシステム構成の
説明図



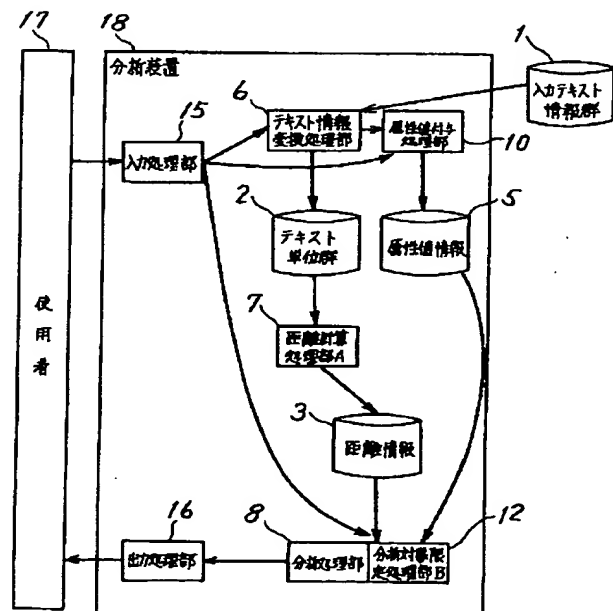
【図 5】

テキスト単位群に名前を付けるシステム構成の
説明図



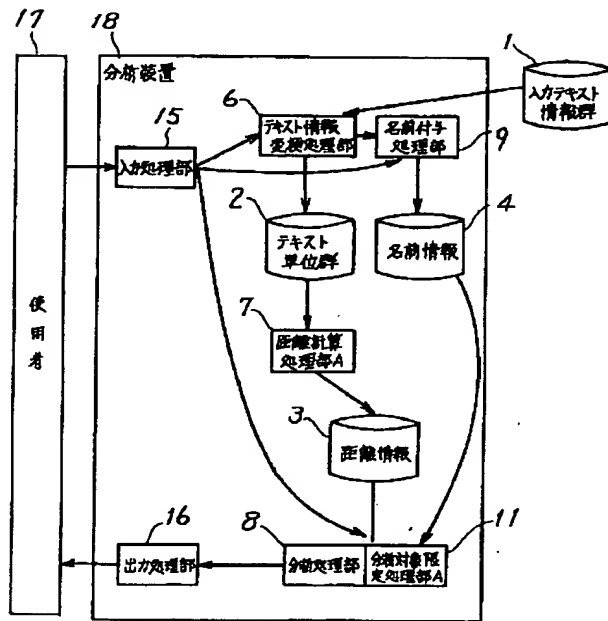
【図 6】

テキスト単位群に属性値を付ける分析装置の
説明図



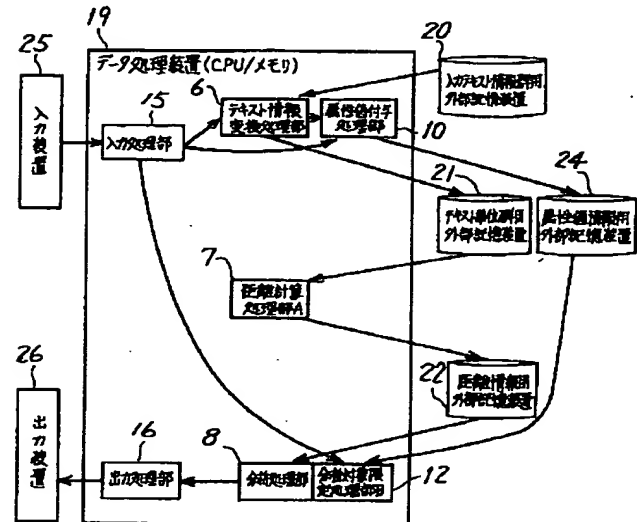
【図 4】

テキスト単位群に名前を付ける分析装置の
説明図



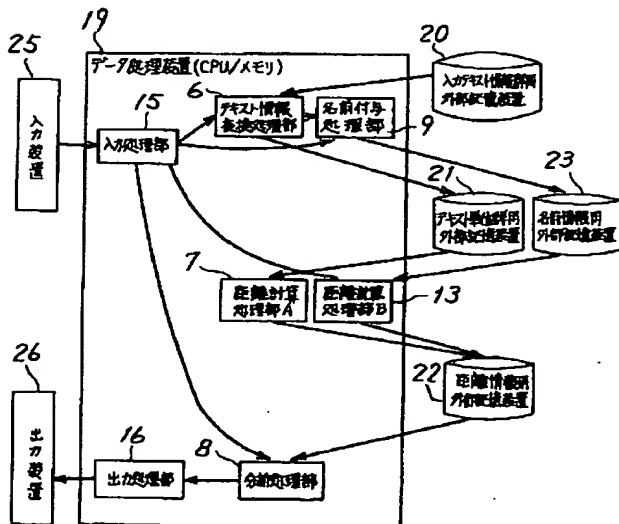
【図 7】

テキスト単位群に属性値を付けるシステム構成
の説明図



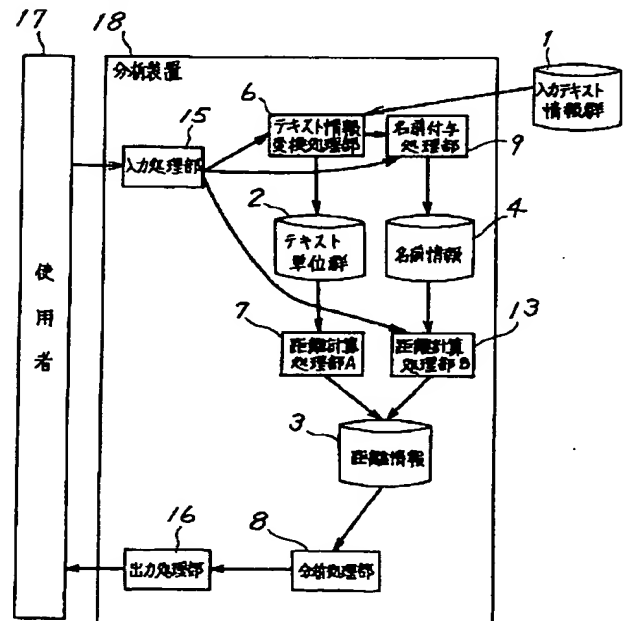
【図 9】

テキスト単位群の名前のパターンから距離を計算
するシステム構成の説明図



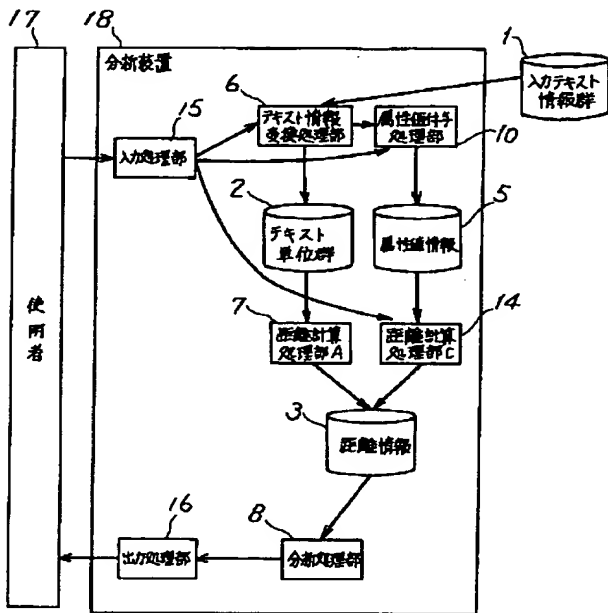
【図 8】

テキスト単位群の名前のパターンから距離を計算
する分析装置の説明図



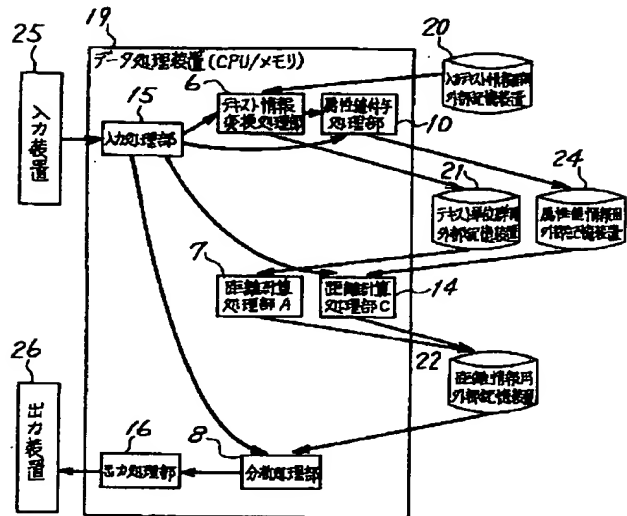
【図 10】

テキスト単位群の属性値のパターンから距離を計算する分析装置の説明図



【図 11】

テキスト単位群の属性値のパターンから距離を計算するシステム構成の説明図



【図 12】

アンケート結果表の説明図

回答者 ID	回答者 の年齢	回答者 の職業	設問 1 : 商品 A のイメージ			設問 2 : 商品 B のイメージ		
			回答欄 1	回答欄 2	回答欄 3	回答欄 1	回答欄 2	回答欄 3
1	35	X	文。文。	文。	文。	文。	文。	文。
2	24	Y	文。	文。文。	文。	文。	文。	文。